DOCUMENT RESUME

ED 100 743 SO 008 023

AUTHOR Taylor, Bob L.
TITLE Potential Uses of the National Assessment Model at the State and Local Levels--Task 4. Final Report.
INSTITUTION National Council for the Social Studies, Washington, D.C.
SPONS AGENCY Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
PUB DATE Sep 73
NOTE 52p.; Related documents are SO 008 019-026

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS *Citizenship; Curriculum Development; *Curriculum Evaluation; Diffusion; Educational Accountability; *Educational Assessment; Educational Objectives; Educational Research; Evaluation; Evaluation Criteria; Measurement Instruments; *Models; National Surveys; Program Effectiveness; *Social Studies
IDENTIFIERS *National Assessment

ABSTRACT

Task 4 of an investigation coordinated by a committee working under the auspices of the National Council for Social Studies describes the model used by the National Assessment for data gathering and reporting on the citizenship area and discusses its potential uses for state and local assessment, curriculum development, and accountability purposes. The seven basic components of the assessment model include objectives development, exercise development, sampling plan, administration of exercises, scoring and analysis, reporting and dissemination, and utilization of information. Maine, Connecticut, Texas, and Colorado have adopted the national model for use in state assessments of learning outcomes. To date three school districts--Lincoln, Nebraska, San Bernardino, California, and Montgomery County, Maryland--are using an adaptation of the model at the district level. Thirteen characteristics for a good assessment program are suggested for developing an adaptation model. The reports on both science and citizenship have resulted in strong recommendations for curriculum changes. As data are gathered at the state level through the use of the model and its adaptations, specific suggestions for changes may be made. Based on the national model, adaptations are proposed which provide a basis for both teacher and program accountability. (DE)

ED 100743

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

POTENTIAL USES OF THE NATIONAL ASSESSMENT MODEL AT THE STATE AND LOCAL LEVELS

TASK IV

FINAL REPORT

Bob L. Taylor

University of Colorado

September, 1973

SP 008 023

This report represents part of an investigation funded by the Education Commission of the States, and coordinated by a special Steering Committee working under the auspices of the National Council for the Social Studies.

TABLE OF CONTENTS

	Page
THE MODEL	2
USE OF THE MODEL AT STATE LEVEL	17
IMPLICATIONS OF THE MODEL FOR CURRICULUM DEVELOPMENT	28
USE OF THE MODEL FOR ACCOUNTABILITY PURPOSES	37
SUMMARY	43
BIBLIOGRAPHY	45

POTENTIAL USES OF THE NATIONAL ASSESSMENT MODEL AT THE STATE AND LOCAL LEVELS

by

Bob L. Taylor
University of Colorado

Under National Assessment, several volumes of data have already been collected and published. There no longer is a debate concerning the pros and cons of having a National Assessment program, for the results of the early assessments are now available to the public. Educators need to study and interpret the results available as a base for decision making in education. Along with this, they need to understand how these data were collected and what were the problems and limitations of the data gathering process.

National Assessment is a plan for the systematic, census-like survey of knowledges, skills, understandings, and attitudes. It is an information gathering plan aimed at providing both educators and the lay public with information concerning the level of achievement in selected subject areas for students and young adults. The goal is to provide information that will be used to improve education. It is concerned with the achievement status of four age levels in ten different subject areas. The subject areas selected for assessment were: Art, Career and Occupational Development, Citizenship, Literature, Mathematics, Music, Reading, Science, Social Studies, and Writing. The basic task of this paper is to describe the model used by National Assessment for data gathering and reporting on the Citizenship area and to discuss the potential uses

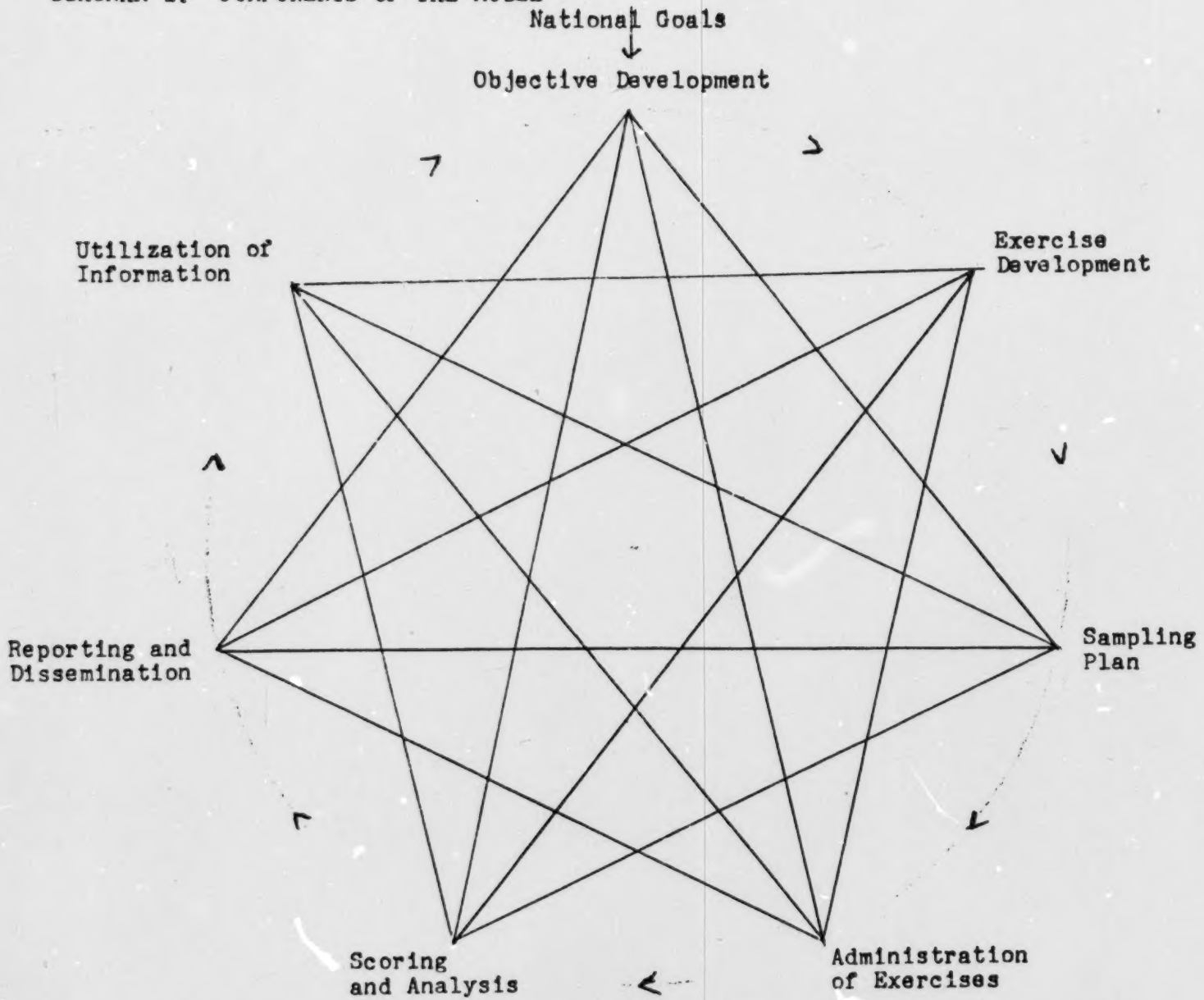
of the model for state and local assessment, curriculum development, and accountability purposes. It is not within the scope of this paper to criticize the model with respect to technical flaws; hence, the model is described and discussed without reference to any potential problems of design.

THE MODEL

The model is in the continuous process of being refined and improved, thus only the basic components of the model are presented in Diagram I. A circular scheme is used in presenting it since, in reality, its actual application may be initiated with any one of the components. Also, in its actual application, there are continual interactions between and among the various components. While theoretically the process starts with the refinement of overall national goals into specific subject matter, behavioral objectives, and progresses in logical sequence through to the final Utilization of the Information, in practice there is much greater freedom with respect to the utilization of the components.

The model for the Citizenship Assessment is presented here in outline form with a fairly detailed description of its components. As presented in Diagram I, there are seven basic components identified in the model: Objectives Development, Exercises Development, Sampling Plan, Administration of Exercises, Scoring and Analysis, Reporting and Dissemination, and Utilization of Information. While many of the fine points of the model are not developed in the following outline, it is described in sufficient detail to give the reader a good understanding of how the data were collected and what

DIAGRAM I. COMPONENTS OF THE MODEL



implications might result from these data. The number of subtopics in the model and their distribution indicate that the major efforts of National Assessment have been with the first five components. The last three components have been areas of controversy and, therefore, have received less attention until recently.

Outline of the Assessment Model for Citizenship[1]

I. Objectives Development[2,3,4,5]

A. The task of developing objectives in the field of citizenship was awarded to the American Institute for Research of Palo Alto, California. These criteria were used in examining the objectives:

(a) They were considered important by scholars.
(b) They were accepted as an educational task by the school.
(c) They were considered desirable by thoughtful lay citizens.

Scholars reviewed the objectives for authenticity with respect to their subject fields; school people reviewed the objectives in terms of their actual emphasis in their schools; and laymen reviewed them in terms of their experiences with regard to their value in life.

B. The American Institute for Research staff reviewed previous lists of citizenship objectives and boiled these down to one comprehensive list of 20 objectives.

---

[1]Womer, Frank B., What Is National Assessment? National Assessment of Educational Progress, Denver, Colo., 1970.

[2]Norris, Eleanor L. (Ed.), Citizenship Objectives, Committee on Assessing the Progress of Education, Ann Arbor, Mich., 1969.

[3]Campbell, Vincent N., et al, Citizenship Objectives for 1974-75 Assessment, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., 1972.

[4]Campbell, Vincent N. and Daryl G. Nichols, "National Assessment of Citizenship Education," Social Education 32:279-81, June, 1969.

[5]Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., November, 1970.

C. Outstanding local teachers familiar with each target-age group (9, 13, 17, adult), working with the American Institute for Research staff, broke down each general objective into the most germane behaviors deemed appropriate as goals for a given age group.

D. A selected group of students and adults in each age group was asked by the American Institute for Research staff to recall and describe outstanding citizens of their acquaintance and specific incidents reflecting good and poor citizenship. These incidents and descriptions, about 1,000, were used to check the completeness of the initial list of objectives.

E. The objectives were stated on three levels (general objectives, sub-objectives, and behavioral age illustrations or statements). The results were summarized for each age group.

F. The revised list of objectives, broken down into important behaviors, was then worked over for three days by a panel of national leaders in citizenship education and related social sciences.

G. A group of persons in various roles from selected California communities reviewed the objectives and made suggestions. These included public and private school administrators, counselors, teachers, a judge, a county planner, labor and business leaders, and social scientists.

H. The objectives were then reviewed by panels of laymen. Eleven lay review panels representing four geographic areas of the country and three different community sizes were used. Each panel spent two days reviewing the objectives based on these two questions: "Is this something important for people to learn today?" and "Is this something I would like to have my children learn?"

II. Exercises Development[6,7,8]

---

[6] Womer, op. cit.

[7] Gadway, Charles J. (Ed.) Reading and Literature: General Information Yearbook, Education Commission of the States, Report 02-GIY, National Assessment of Educational Progress, Denver, Colo., May, 1972.

[8] Finley, Carmen J. and Frances S. Berdie, The National Assessment Approach to Exercise Development, National Assessment of Educational Progress, Ann Arbor, Mich., 1970.

A. The production of the exercises was initiated by the American Institute for Research in 1966. The exercises were developed to cover all of the major objectives and to represent the selected content areas. Many exercises required the use of interview techniques, as well as the usual pencil and paper exercises. Also, self report and group task exercises were used.

B. Because National Assessment intends to describe what people in an age group know, the exercises were written to reflect three difficulty levels--reporting knowledge or skills common to almost all persons in an age group, reporting skills or understandings of a typical member of an age group, and reporting understandings or knowledge developed by the most able persons in an age group.

C. All exercises were developed to meet these criteria: content validity, clarity, functional exercise format, clustering exercises based on a single set of stimulus materials, directionality of response, difficulty level, content sampling, and overlap between age groups. The exercises were direct measures of some pieces of knowledge, understandings, attitudes, or skills which were mentioned in one or more of the objectives.

D. The exercises were reviewed by panels of lay persons for clarity, meaningfulness, and invasion of privacy.

E. There was a tryout of the exercises involving representatives of groups in the actual assessment--regions, communities, races, sexes, and age groups. Following the tryouts, the American Institute for Research staff and subject matter specialists reviewed the tryout data and made needed revisions.

F. A committee of subject matter specialists, measurement specialists, and National Assessment staff members rated the exercises to be included in the packages according to a set of criteria, and based on the ratings the exercises were selected.

G. The selected exercises were reviewed by U. S. Office of Education personnel for any infringement of privacy on the part of the respondents or possible offensiveness.

H. Since there were about 160 minutes of testing time available for each age group in each subject area, the exercises used were only a small sample of the potential number of exercises. The exercises were assembled into administrative units (packages) for groups up to 12 persons.

III. Sampling Plan[9,10]

A. The sampling plan was subcontracted to Research Triangle Institute, Raleigh, North Carolina. A multi-stage design was used which was stratified by region, size of community, and socio-economic status. This was a probability sample which allowed researchers to collect data from a small sample of the population and to infer from that sample certain characteristics of the entire population.

B. The populations for assessment were all 9 year olds, all 13 year olds, all 17 year olds, and all young adults 26 through 35 years old in the 50 States plus the District of Columbia. The only exceptions were the exclusions of institutionalized individuals of these ages--those in hospitals, prisons, and others who could not be reached.

C. For ages 9 and 13, a school sample only was used and for the 26 through 35 age group a household sample only was used. For the 17 year olds, both a school and a household sample were used.

D. The entire country was divided into population areas as follows: cities, counties exclusive of cities, and pseudo-counties--two or more counties were put together when the population of a single county was less than 16,000. Each population unit of 16,000 residents was assigned a number.

E. The country also was divided into four geographic regions: Northeast, Southeast, Central, and West.

F. Each geographic region was divided into communities of four types: large cities of above 180,000 population, urban fringe, middle-sized cities between 25,000 to 180,000 population, and small town-rural of under 25,000 population.

G. The 52 sampling units for each geographic area were spread across the four community types in a fashion proportional to their population in relation to the area population.

H. To insure comparable representation from each part of the country, an equal number of sampling units was selected from each geographic region--52 from each of the four regions for a total of 208.

---

[9]Norris, Eleanor L., et al, Report 1, 1969-1970 Science: National Results and Illustrations of Group Comparisons, J. R. Chrony and D. G. Horvitz, "Structure of Sampling and Weighting," Appendix C, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., July, 1970.

[10]Norris, Citizenship Objectives, op. cit.

I. The actual sampling units chosen were selected at random. This plan did not guarantee that all 50 States would be included in the sample. This was not a survey objective, but later the design was changed so each state was included in the sample.

J. In each sampling unit selected, all school buildings enrolling students of the sample ages (public, private, and parochial) were identified.

K. The plan for schools was to select units of approximately 250 to 350 pupils for each age group and from at least two different buildings within each sampling unit for each age group.

L. Each cooperating building principal provided a list of names of students in the building from the specific age groups. This list was used for the final random selection of students to take the assessment exercises from that building.

M. Information about the areas was obtained from the U. S. census data. In order to report reliable information for lower socio-economic status groups, these groups were oversampled. There was a disproportionate number of schools from lower socio-economic status areas included. In the overall results, the data from the lower socio-economic areas were given the percentage value in which they occurred in the total population.

N. From each of the 208 geographical samples, 100 adults, ages 26 through 35, were randomly selected using the following procedures. Each of the 208 geographic samples was divided into equal secondary sampling units. Then ten secondary sampling units were randomly selected from the total 208 samples. Interviewers then personally contacted the people in the chosen secondary sampling units of the 26 through 35 age group and out-of-school 17 year olds. These persons were asked to participate in the assessment.

O. Individuals were classified as black, white, and other on the basis of information provided by the school or by observation. Results were given for black and white only. The number of individuals classified as other was too small to produce reliable results.

IV. Administration of Exercises[11,12]

---

[11] Womer, op. cit.

[12] Gadway, op. cit.

A. The administration of the exercises was subcontracted to Research Triangle Institute in the East and to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa, in the West. Cooperation of schools was obtained by first contacting officials at the state and then at school district levels. There was above 90 per cent cooperation by schools.

Adults and out-of-school 17 year olds were contacted by a personal door-to-door household canvass. Each out-of-school participant was contacted individually. All had the right to refuse to cooperate.

B. A full-time trained staff of 27 district supervisors managed the field work. They were assigned to different geographical areas of the United States. They contacted schools and recruited and trained local teachers to help in the administration of the exercises in schools and recruited and trained other available persons for the out-of-school administration.

C. In the schools, students from a single age group from different classes were brought together in a room for exercise administration. Group size was at least 8 and usually 12 students.

D. The exercises were organized in packages which contained exercises from two or three different subject areas at a single age level. No one person took all the exercises in his age group. Age groups were assessed at different times of the year.

E. In packages administered to groups, taped directions and taped readings of the exercises were used in addition to printed packages. This was done to establish consistency in timing and administration plus to provide for non readers.

F. Several packages at ages 9, 13, and 17 consisted of exercises that were given by exercise administrators to one individual at a time. The administration of all the packages for the adult assessment was done by interviews.

G. Each package required about 50 minutes of administrative time. Each person took only one package with the exception of the out-of-school, 17 year olds who were asked to take four or five packages each since they were the most difficult and expensive group to locate.

H. Students' names were confidential and did not appear on any packages. The name roster was kept at the building level and used only in the organization of the in-school sampling.

V. Scoring and Analysis[13,14]

---

[13]Womer, op. cit.

[14]Gadway, op. cit.

A. The scoring and analysis of the exercises were subcontracted to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa.

B. The multiple-choice exercises were scored and recorded routinely by machine.

C. The openended exercises were scored by trained professionals using a key of acceptable and unacceptable achievements in terms of the objectives.

D. Results were reported for each goal. Also, the results were reported both as the percentage of any group of respondents making the desired responses to an exercise and as the difference between the percentage of a group making the desired responses and the corresponding national percentage.

E. In the assessment, there was a lack of proportionality among characteristics used in the comparison of groups, such as color, sex, parental education. A statistical procedure, balancing, was used to correct for this problem in the comparative analysis of the data. Balancing is a procedure to examine the performance of groups classified on one characteristic adjusting for the fact that these groups differ on a specified set of other characteristics.

VI. Reporting and Dissemination[15,16,17]

A. The reporting of results was directed to subject matter specialists, professional educators, and informed laymen. Multiple reports were developed to serve these different audiences.

B. Approximately 40 per cent of the exercises were reported at the end of each assessment year. Not all exercises were reported since they were to be used over again in future assessments in order to measure change by means of comparing the results on the uncontaminated exercises.

C. The exercises released for publication were selected to be representative of all exercises administered as well as the results received on the assessment.

---

[15] Womer, op. cit.

[16] Campbell, Report 2, Citizenship: National Results, op. cit.

[17] Gadway, op. cit.

D. Reporting was done by 9, 13, 17, and 26 through 35 age groups. Since the same exercises were used with different age groups, there was comparable data across two or more age levels.

E. Reporting was also done by groups within the categories of regions, community types, sex, socio-economic status, and white, black, and other.

F. Final reports were printed with a short description of the exercises, the national percentage of success, and group differences from the national percentage of success for each exercise. This was done without any interpretation of results.

G. Both observed and balanced results for all exercises and by groups were reported. The effects of balancing on measured characteristics such as sex and region were included in the report.

H. There were no scores reported for individuals. No single individual took more than one twelfth of the exercises, and no individual took a package that sampled a single subject area.

I. Results were reported through the media: written word, radio, television, films, and personal reports.

VII. Utilization of Information[18,19]

A. The results provided information for educational decision making. For example, considering the somewhat lower performance of the Southeast Region on the Citizenship results, school boards in that region might decide to put greater stress in their school programs on citizenship skills, understandings, and attitudes.[20]

---

[18] Womer, op. cit.

[19] Conaway, Larry E., "Some Implications of the National Assessment Model and Data for State and Local Education," Paper Presented at the 1973 Annual Meeting of the AERA, New Orleans, February 26, 1973.

[20] Campbell, Vincent N., Manford J. Ferris, and Daryl G. Nichols, National Assessment Report 6, 1969-1970 Citizenship: Group Results for Sex, Region, and Size of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, July, 1971.

B. The results raised many questions which may lead to other investigations. For example, in making comparisons of all Citizenship results combined, it was found that the Extreme Affluent Suburbs showed substantial median advantages at all ages and that the Extreme Rural and Extreme Inner City showed substantial deficits at all ages. Here are discrepancies in performance which need to have causal studies conducted on them from the perspectives of different disciplines such as political science, sociology, economics, and education.[21] A second example is that assessment could expand into other educational areas, such as assess a sample of 5 year olds for a basis of comparison, break results down by states, or expand into new subject areas.

C. The results of several cycles should provide evidence of the change in knowledge, skills, understandings, and attitudes in the age groups as they relate to educational objectives.

D. School administrators can make comparisons between groups, and may improve student performance from the information gained in this manner.

From this review of the model, it is evident that the National Assessment staff has put a great deal of effort and know-how into the design, plus the development of each of the components. Each of the components reflects the best available information in the field of assessment.

A complicated but logical process has evolved for identifying the behavioral objectives for the specific subject areas. The use of laymen along with teachers and subject-matter specialists added validity checks on these objectives. Considerable inventive talent was displayed in writing the exercises for assessing some of the non-cognitive type of objectives. The use of the tape-paced administration

---

[21] Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size, and Type of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, May, 1972.

was a valuable innovation in the field of testing. In this procedure, a complete set of directions for the administration of a package of exercises is recorded on an audio tape. By playing the tape, the directions are always given in the same manner and the exercises are timed. Also, the problem of nonreaders is overcome through the oral directions. While the technology had been available prior to this use, it had not been widely used. The procedures for scoring subjective exercises have provided a model for accurate, objective scoring of these kinds of exercises. The scorers were put through a special training program for scoring essays, written materials, or performance exercises.

The results have been released in census-like fashion without interpretation. Clear, factual reports were made of the data so the reader would know what was assessed and what were the results. Very little interpretation of the results has been done for the reader. This was the intent of the National Assessment staff, for they did not conceive of their role as going beyond the reporting function.

The final two components of the model are the most controversial, for originally there was not a clear mandate for National Assessment to take a vigorous leadership role with respect to these two tasks. As time has gone by, greater activity by National Assessment in these areas has been urged by such groups as educators and parents, and National Assessment has become more active in the dissemination and utilization components.

In summary, it can be pointed out that the model reflects some important choices on the part of the staff of National

Assessment. The decision was made to assess a broad range of objectives in each of the ten subject-matter fields. Certainly, it would have been much easier and cheaper to have concentrated on a narrow set of objectives. Also, the effort was successfully made to include the higher cognitive levels in the assessment exercises and to deal with the affective domain. In subject areas like Citizenship and Social Studies, the usual standardized test concentrates on factual knowledge which is of a less controversial nature and easier to assess. National Assessment should be commended for their bolder, more comprehensive approach to the task which searches for more significant kinds of data.

Changes in the Model

From the time the assessment in Citizenship was carried out until the Social Studies assessment was conducted, there were changes made in the assessment model. This can be illustrated by the changes made in the procedures for identifying objectives. For the Citizenship objectives, the American Institute for Research at Palo Alto, California was the contractor, while for the Social Studies objectives, the contractor was Educational Testing Service, Princeton, New Jersey. Referring to the Objective Section of the Outline of the Model, you can compare the process used in establishing the Citizenship objectives with those presented here for establishing the objectives for the Social Studies.[22]

---

[22] Norris, Eleanor L. and Barbara Goodwin, (Eds.), Social Studies Objectives, Education Commission of the States, National Assessment of Educational Progress, Denver, Colo., 1970.

The following were the steps in the procedure used in establishing the Social Studies objectives:

1. The task of developing objectives in the Social Studies was awarded to Educational Testing Service, Princeton, New Jersey.

2. A committee of 11 social scientists plus Educational Testing Service staff members prepared an initial draft of the objectives for Social Studies.

3. From the nominations by officers of various national and state organizations in education, persons were selected to form 11 lay-review panels. These individuals represented rural, small town, suburban, and large cities in the four geographic regions of the United States.

4. A social studies conference of 21 people including scholars from the fields that contribute to the area of the social studies, leading social studies teachers, university professors who specialize in social studies education, and representatives of research organizations with potential for reworking the objectives met to discuss the need for revision and for clarification of the existing statement of objectives.

5. Following the design used for the development of the original objectives, the Educational Testing Service staff prepared a preliminary revision of the list of objectives.

6. A committee of four social studies educators met and reviewed the Educational Testing Service staff's preliminary revision.

7. The revised statement of objectives was submitted to additional reviews, one by subject matter specialists and the other by elementary and high school teachers.

8. The objectives were then mailed for reactions to one member of each of the 11 lay panels which had reviewed the objectives earlier. The revised report was then accepted as the statement of objectives for the Social Studies assessment.

There are important differences in how the Citizenship and Social Studies objectives were developed. To some extent, this reflects the respective characteristics of the two fields. In the development of the Citizenship objectives, greater stress was placed on input from laymen, students, and teachers. Also, while the social studies effort put more stress on academic representation, it also went for more widespread geographic representation.

The Citizenship effort sought wider representation from different groups of people, but it was concentrated rather heavily in California.

The operational details of the components will vary from one assessment to the next depending on such factors as what group is awarded the contract for different parts of the assessment, what is the subject field being assessed, or what academic scholars are involved in the assessment. This flexibility makes for both strengths and weaknesses. A strength is that there are variations in the approach and old problems may be corrected. Weaknesses are that different assessment cycles may reflect varying philosophic positions both within the subject field and in the assessment process so that the data may not be strictly comparable. Also, it does mean that anyone studying assessment results must determine just how the data with which he is concerned were collected, for there may be important differences in the assessment process with which he must be familiar in order to interpret the data correctly.

The above descriptions demonstrate that revisions in the assessment process are ongoing and that the model is constantly being changed with respect to the details of the components. While it is possible to identify the major components of the model, it is not possible to give a precise, up-to-date description of the components because they are in the process of continual change. This is one reason why the Citizenship Assessment model was used in this paper, since it (with hindsight) could be reported in greater detail, and it does provide the reader with a good understanding of how the National Assessment process operates.

USE OF THE MODEL AT STATE LEVEL

National Assessment is a census-like study to collect information concerning the educational attainments of Americans. In planning for the collection of this census-like data, the model, which was presented in the previous section, was developed. A number of states have found adaptations of the model useful in conducting state assessments in which desirable learning outcomes are identified and the status of learners with respect to these outcomes is determined.

State assessment is a rapidly developing movement. At this writing, all of the states have assessment activities either in operation, in a developmental process, or in a planning stage.[23] While the statewide assessment programs have many similarities, they break down into two basic types of programs on the question, "Who gets to use the results?" The divisions are those states for which data are collected for decision making by state agencies and those states for which data are collected for decision making by teachers and administrators. State programs for which the emphasis is on collecting information for state level decision making are: Arizona, California, Colorado, Connecticut, District of Columbia, Florida, Maine, Massachusetts, Michigan, Nevada, New Jersey, New York, North Carolina, Rhode Island, South Carolina, Tennessee, and Texas. Programs for which the emphasis is on collecting information

---

[23] State Educational Assessment Programs, 1973 Revision, Joan S. Beers and Paul B. Campbell, "Statewide Educational Assessment," Educational Testing Service, Princeton, New Jersey, 1973, p. 1.

for local level decision making are: Alabama, Arkansas, Delaware, Hawaii, Idaho, Iowa, Kentucky, Mississippi, New Hampshire, New York, North Dakota, New Mexico, and Pennsylvania.[24]

Beers and Campbell report that a number of characteristics are appearing in these assessment programs. In about a third of the states, the programs were mandated by the state legislatures, and the results of the assessments are to be reported back to the state legislatures. In a few of the states, the data are to be used for PPBS (Planning, Programming, and Budgeting Systems). In about half of the states where the assessment data are being used to make state-level decisions, state and Federal funds will be allocated based on the results. Participation in assessment is required by law in about a fifth of the states. In the states where the assessment data are being used to make state-level decisions, samples rather than all students are being assessed, while in the local-level, decision-making states all students in the target populations are being assessed. Criterion-referenced instruments are very common with the states where the data are being used for state-level decisions, but the states collecting information for local decision making are favoring norm-referenced instruments. Finally, no dominant funding pattern has evolved in either of the two groups of states.[25]

State Adaptations of the Model

In the assessment of Citizenship education, Maine made an extensive application of the National Assessment model and carefully

---

[24] Ibid., p. 1.

[25] Ibid., pp. 2-3.

duplicated it so that comparable data were collected at the state level.[26] Maine's first cycle of the ten subject-matter areas of National Assessment is to be completed by scheduling two of these areas each year for five years. Citizenship and Writing were the first subject areas to be assessed.

Based on the results of a previous study of objectives for education in Maine, two review committees decided to accept the National Assessment objectives as being closely related to the Maine objectives. Maine selected the 17-year-old population of in-school students for its first assessment. A sample of 2,000 17-year-old students was used to represent the approximately 17,000 17-year-old students in the State. The State was divided into four geographical regions. As in National Assessment, school buildings were randomly selected from the geographic regions, and students were then randomly selected from the buildings. Packages were developed with exercises taken from the two subject areas. The available, released exercises from National Assessment were carefully examined to see if they reflected objectives valid for Maine and to see if some could be modified, where needed, to be administered in group sessions using the paced-tape method while still retaining a high degree of comparability to the National Assessment individually administered exercises. The packages were made up of 23 Citizenship and seven Writing exercises, plus a 23-item Student Questionnaire. The exercise format was kept virtually identical to the one used in National Assessment. Trained

---

[26] Maine Assessment of Educational Progress: Methodology (Report 5), Department of Educational and Cultural Services, Augusta, Maine, 1972.

administrators were sent out to administer the exercises, and the exercises were scored according to National Assessment procedures. On data reporting and analysis, there was the census-like reporting of the performance of the Maine students plus comparisons of the Maine results with appropriate National Assessment data.

In summary, the Maine Assessment duplicated the National Assessment procedure as completely as possible. With minor exceptions, the same objectives were used for Citizenship. The same sampling design was used with adaptations to a smaller geographical area and population. The exercises were for the most part taken from those released by National Assessment, and they were organized into packages similar to those used by National Assessment. The administration and scoring of the exercises were conducted in the same manner as National Assessment worked. Since the same private contractors were used by Maine as were used by National Assessment, the duplication was complete wherever possible. The reporting and data analysis were similar, and the data did provide the opportunity to compare the results in Maine with the results from National Assessment.

Here, the model was very carefully duplicated at the state level. The big question which comes to mind after studying the Maine Citizenship report is, "Aren't the National Assessment data being treated here as some kind of a national norm against which the performances of 17-year-old students in Maine were being compared?" Of course, this use of National Assessment data had been questioned from the start of the proposal for an assessment at the national level. Now, Maine has provided the opportunity to study

the effects of this use of the data on the educational system of a state.

Another state which carefully followed the model was Connecticut.[27] Here, an assessment was first conducted in Reading. To permit comparisons, the Connecticut program used available instruments and applicable procedures developed by National Assessment which were adapted to the requirements of the local situation. Connecticut's Reading objectives were matched to the Reading objectives of National Assessment. Approximately 220 reading exercises from National Assessment were used in producing the packages used in the Connecticut assessment. Exercises were selected to represent all of Connecticut's Reading objectives. The age groups assessed were 9, 13, and 17. As with the National Assessment packages, tape recorded instructions were used. The sampling design was a multi-staged design duplicating with few exceptions the National Assessment design. As with National Assessment, a group of administrators for the packages was recruited and trained.

This was another example of careful duplication of the National Assessment model down to using the same objectives and exercises. Again, there was the use of the National Assessment results as norms to which the Connecticut results were compared.

The Texas Needs Assessment used the model for the development of their assessment in mathematics at the sixth-grade level.[28]

---

[27] Report on the Assessment of Reading Skills of Connecticut Public School Students, Institute for the Study of Inquiring Systems, Philadelphia, Pa., and Department of Education, Hartford, Conn., 1972.

[28] Sixth-Grade Mathematics: A Needs Assessment Report, Texas Education Agency, Austin, Texas, 1972.

However, while using ideas from the model, they broke with it in a number of places. The Texas people were concerned that the assessment would yield information which would be useful to teachers in their classroom instruction of students. From a pilot study, it was decided to use a criterion-referenced reading test and to work with grade levels instead of age groups of students. They worked with the sixth grade, and the tests were administered by the staff of each school which participated in the assessment. The objectives were chosen from the major skill areas treated in the state-adopted textbooks. Regional location and community size were taken into consideration in selecting the sample. Approximately 10 per cent of the Texas schools teaching at the sixth-grade level administered tests, and approximately 10 per cent of the pupils being taught at the sixth-grade level were included in the sample. Reports were given to teachers on the performance of their individual students. Also, there was a school report on the performance of the students for each school and a report on each of the classes in the school. Comparisons were made on the basis of sex, race, and size of community.

The Colorado Needs Assessment, while using the model, made an even greater break with it.[29] Its objectives were based on a state study of educational goals, and the educational goals were restated in terms of performance objectives. Following the model, objective-referenced exercises were written. A sampling design was used and the student responses were analyzed. In this assessment, classroom teachers were involved in the writing and refinement

---

[29]Helper, John W., An Assessment of Learner Needs in Colorado, Colorado Department of Education, Denver, Colo., 1972.

of the behavioral objectives. Objective-referenced exercises were written for nine subject areas. The exercises were administered to a sample of 30,000 Colorado students. A stratified random sampling procedure was used to select a sample of school districts of the State. Then schools were selected at random from the districts chosen. Finally, classes in school buildings were randomly chosen for testing. The samples were representative of all Colorado students in grades 3, 6, 9, and 12. A group of proctors was hired and trained to administer the exercises, and the exercises were scored by computer. The data were analyzed on a statewide and district basis, and the results were broken down by subgroups, e.g. boys, girls, urban, rural.

As pointed out earlier, states are rapidly moving into the assessment field. Some are reproducing the National Assessment model at the state level, and others are developing variations of the model. The more crude efforts have resulted in endless pages of raw percentages without any explanation of the results. Based on a survey of state assessment programs, Beers and Campbell identified several of the problems which are common to these state programs.[30] Naturally, a shortage of money and staff were the most frequently mentioned problems, for it is a fact that many states have moved into this area without providing adequate funds for a realistic assessment program. Also, teacher resistance to assessment and negative public attitude toward outside testing were problems mentioned. Test results have been misused in the past, such as the firing of teachers based on incorrect interpretation of test results. Also, test developers have been guilty of violating the privacy of students through questions which

---

[30] State Educational Assessment Programs, Beers and Campbell, op. cit., p. 3.

transgressed the examinee's human and legal rights. A third problem area has been with the utilization and dissemination of results. Some school officials do not understand the results. In some situations, there has been hostility to the results. Some officials have ignored results in making decisions. Finally, results have frequently not reached the right people in a useable form.

Use of the Model at the District Level

To date, there have been a limited number of efforts reported on the use of the National Assessment model at the district level. Three such assessments on which some data have been released are being conducted in Lincoln, Nebraska; San Bernardino, California; and Montgomery County, Maryland.

In Lincoln, Nebraska, the exercises released by National Assessment in Citizenship and Writing were used in a local assessment which yielded data comparable to National Assessment data.[31,32] A group of supervisory personnel from the central office identified the Citizenship objectives which were applicable to the Lincoln schools. Then the released National Assessment exercises were selected which were applicable to these Lincoln objectives. Also, the National Assessment model was followed in selecting a random sample of 13 year olds from the Lincoln junior high schools. In addition, a sample of in-school, 17 year olds was tested on some

---

[31] "Weekly Focus," Lincoln Public Schools, Lincoln, Nebraska, February 12-19, 1973, p. 3.

[32] Brandt, Ronald, Associate Superintendent for Instruction, Report on Assessment Results to Board of Education, Lincoln Public Schools, Lincoln, Nebraska, Spring, 1973.

of the writing exercises. The administration of the exercises was carried out by a group of specially trained administrators, and the tape-paced method was used in presenting the exercises to the students. Scoring followed the National Assessment procedures, and in reporting the results comparisons were made to National Assessment data with special attention given to comparable subgroups such as cities of similar size and the same geographic region.

The San Bernardino City Schools developed a criterion-referenced assessment model of student progress which was based on the National Assessment model.[33,34] This model involved local teachers, students, and laymen in a good setting; eight educational goals were identified through the efforts of workshops involving teachers, students, and patrons. A Curriculum Task Force composed of 20 teachers wrote behavioral objectives for the goals to be appropriate for grades 3, 5, 9, and 12. National Assessment consultants assisted the teachers in developing exercises to assess the stated objectives at these grade levels. Also, the National Assessment consultants helped to design a sampling procedure to provide district-wide representation. The exercises were organized into test batteries for each grade level. The Teacher Task Force administered and scored the tests. This is a break from the National Assessment

---

[33]Bonney, Lewis A., "Application of the National Assessment of Educational Progress Philosophy in San Bernardino City Unified School District," Unpublished Paper, San Bernardino City Unified School District, San Bernardino, California.

[34]Special Curriculum Task Force, "Report on Student Performance," Office of Instructional Services and Research and Development Office, San Bernardino City Unified School District, San Bernardino, California, June, 1972.

practice of using specially trained exercise administrators. The results were tabulated in terms of percentage of students meeting stated behavioral objectives. While the National Assessment model was followed in many ways, such as use of behavioral objectives, criterion-referenced assessment instruments, and sampling of target populations, the assessment was designed for application at the local level, and it was planned, developed, and carried out by local personnel.

The Montgomery County Schools, Maryland, developed a program for assessing 13- and 17-year-old students.[35] In this assessment, the released National Assessment exercises for Writing were used. These were administered in two group-package sessions to samples of 13- and 17-year-old students. The results for Montgomery County students were compared to the results from the nationwide samplings of 13 and 17 year olds by National Assessment. One of the variations in the Montgomery County sampling design was stratification by I.Q. and grade level. The purpose was to spread the sample across the grade by school-I.Q. groups; however, these sampling groups were not used as reporting units. Each age group (13 and 17) was stratified by I.Q. groups (low and non low) and by grade levels.

Adaptation of the Model

In the above discussion, it is evident that there will be as many adaptations of the model as there are local and state units

---

[35] Bayless, David L., Ralph E. Folsom, and Louise H. Lewis, "Sample Design for Assessing Montgomery County Public Schools 13- and 17-Year-Old Pupils Using the NAEP Model," National Assessment of Educational Progress and Educational Commission of the States, Denver, Colorado, January, 1973.

conducting assessments. Probably, there is no specific assessment model which is the best; hence, there is no model that should be applied without modification in any and all situations. Nevertheless, there are principles of good assessment which should be applied in developing or adapting a model for local assessment purposes. Listed below are some characteristics which should be found in a good assessment program.[36]

1. The program has clearly defined goals that apply to a particular audience or audiences.

2. The program has a realistic number of goals which are attainable under the existing assessing conditions.

3. The program has established priorities among its goals and places its major efforts on its major goals.

4. The program has been designed to gather information considered to be important in education.

5. The program has specific objectives which it is striving to attain.

6. The program has been designed to provide results at a useable level of accuracy.

7. The program has used data-gathering instruments which measure the objectives of the assessment.

8. The program has collected data in such a manner as to introduce a minimum of error in the results.

9. The program has scored and processed data in an accurate manner.

10. The program has used analytic techniques that provide the data breakdowns needed by decision makers.

11. The program has reported results in a manner useable by its audience.

12. The program has provided help in the interpretation of results and assistance in their implementation.

13. The program has provided for the active involvement of groups of persons from all of the major audiences for the assessment results.

---

[36]Womer, Frank B., Developing a Large Scale Assessment Program, Cooperative Accountability Project, Denver, 1973, p. 89.

IMPLICATIONS OF THE MODEL FOR CURRICULUM DEVELOPMENT

Of course, a major, potential outcome of National Assessment and the model was the providing of new, accurate data with regard to curriculum problems. Curriculum decision makers are furnished data which have not been available to them before this. Because of this new information, they should gain new insights into their problems, and hopefully there will be innovative approaches taken to the solution of these problems resulting from this.[37]

Implications of the Model

The assessment model has potential for promoting curriculum development. This is especially true when it is applied to state or local situations in the manner used in Colorado and San Bernardino. In these two situations, objectives were developed which specifically applied to the local situation. The statement of well written objectives in behavioral terms may sharpen the purposes of instruction. Through the experience of writing behavioral objectives, the curriculum worker gains a much clearer perception of his task; hence, this practice may have a beneficial impact on curriculum work. On the other hand, the use of behavioral objectives has not always been a positive influence. The objectives may zero in on easily defined behaviors which lack scope and significance. They may produce tunnel vision, and put stress on the inconsequential and trivial. In an effort to be specific and to define the exact behaviors desired, the larger perspective may be lost.

---

[37] Womer, Frank B. and Marjorie M. Mastie, "How Will National Assessment Change American Education?" Phi Delta Kappan 53: 118-20, October, 1971.

Again, the development of exercises from the identified behavioral objectives may have a positive influence on curriculum. The kind of new, innovative exercises which have been developed by National Assessment may have a very positive influence on what is being taught and how it is being taught. Teachers both in reviewing exercises which have been used in National Assessment and in writing exercises for local assessments may be influenced in their selection of both content and methods by their knowledge of these assessment exercises. Material not relevant to the objectives of the course may be dropped, and methodologies promoting the kind of skills needed in the assessment exercises may be introduced.

On the other hand, the results may be less desirable. If in local and state situations the dictates of finances or the lack of leadership result in the use of poorly written, machine-scored, multiple-choice exercises, the results may be very negative. Teachers may feel pressured to stress rote learning of facts in order to prepare their students for poorly written examinations. Hence, poorly written exercises may keep irrelevant material in the curriculum and limit curriculum innovation and development. The quality of the exercises written and released will have an impact on curriculum development.

Good sampling procedures may give insight into the status of knowledge, understandings, skills, and attitudes of students in a particular target population. This can promote curriculum improvement and innovation. Problem areas in the curriculum may be identified. From the National Assessment, there have been some problem areas identified in the Citizenship results. On an exercise

dealing with freedom of speech, a large percentage of 13, 17, and adult age groups indicated that they would not allow sample controversial statements to be made on radio or TV.[38] This showed a lack of understanding or valuing of the Constitutional right of freedom to experss controversial or unpopular opinions.

The results on the Citizenship assessment indicated that black, urban students in our large cities compared poorly on knowledge about the structure and function of government to the national average performance on the same exercises.[39]

On the other hand, there are potential difficulties with assessment data which represent national levels of performance. Even though the data were not collected with this intention and were reported in census-like form, the results of National Assessment are being treated like national norms. Several states have conducted their own assessments duplicating the National Assessment model so that they can make direct comparisons between their state results and the various national, regional, and subgroup results. There is the potential of great mischief in this approach, for it may lead to unfair comparisons between groups, states, and regions. In the assessment reports of some states, tables of percentages have been presented without any interpretation or explanation. Some school systems have been presented in a very bad way without any reference being made to the kinds of variables involved in the different learning situations. Such variables as per pupil

---

[38]Campbell, Report 2, Citizenship: National Results, op. cit., p. 34.

[39]Norris, et al, National Assessment Report 9, op. cit., pp. 63-65.

expenditures, educational level of parents, and motivation of pupils do have an impact on the learning situation. These and other variables cannot be ignored in interpreting the results of assessment.

Here, it is not being suggested that assessments should not be conducted because there are potential misuses of the data, but it is being pointed out that misuses of data do occur. This does mean that those engaged in assessment at national, state, and local levels have the responsibility to be constantly engaged in an educational program to aid those using the data to make correct interpretations of it. We need these kinds of information for decision making, but if the data are misused or misinterpreted, then the decisions based on them may not be good ones.

Impact on Curriculum

National Assessment is providing data on which decisions can be made. The reports on both Science and Citizenship have resulted in strong recommendations for curriculum changes in these subject areas.[40] As data are gathered at the state level through the use of the model and its adaptations, specific suggestions for changes may be made. For example, Texas has designed its state assessment so that there is direct feedback at the classroom level. This may have a strong and immediate impact on these classrooms which could be either good or bad depending on what interpretations and recommendations are made.

---

[40] Harrison, Charles H., "Are We Educating for Tomorrow?" Scholastic Teacher, September 21, 1970, pp. 16-19.

There are other problems which should be pointed out concerning curriculum decision making and national and state assessment results. While the National Assessment process for identifying objectives provides for broad-based participation in the decision-making groups, objectives still are selected which neglect sizeable, subcultural groups in our nation. This means that even some of the very general objectives which are selected may not apply to these groups, and thus they are not valid for some situations.[41]

To illustrate the potential difficulty with objectives and subgroups, consider Objective V of the Social Studies. This is, "Have a reasoned commitment to the values that sustain a free society." This has under it subpoint B, "Believe in the rule of law and can justify their belief." For blacks in the core city, this objective may not carry value. Their experiences may have been that the law is used against them, and they need to constantly be looking for new ways to work around the law.[42] While this is perhaps an extreme example, it does make the point that in as diverse a population as we have in the United States, there will be many subgroups for whom the broad general objectives do not carry the same meaning as for other members of the nation.

There is the question of what are the components of a general education or what should be the components. The selecting of

---

[41]Wrightstone, J. Wayne, Thomas P. Hogan, and Muriel M. Abbott, "Accountability in Education and Measurement Problems," Test Service Notebook 33, Harcourt, Brace, Javanovich, Inc., New York, p. 3.

[42]Norris and Goodwin, Social Studies Objectives, op. cit., p. 26.

common objectives for a subject area such as Science and the writing of exercises for these objectives provide one definition of what students of a certain age group are expected to know about science. Since this establishes what comprises general education in science, it has a definite limiting impact on the freedom of choice of a student in deciding what he wants to learn. He is being dictated to concerning what he should take in general education. This situation always has existed in education, but it puts the assessment movement in the camp of the conservatives in the current controversy with respect to free choice and unlimited electives for students. Here, the assessment movement is counter to the humanizing movement in American education.[43] It is promoting a closed rather than an open approach to curriculum.

The very identification of a subject area for National Assessment will have an impact on the fortunes of that subject area. The "chosen" subject areas are more likely to be supported financially and retained in the curriculum than those not selected by National Assessment. Those subject areas which are not included in the "magic" ten may find that they are second class so far as school boards are concerned. If the students of the district have made a poor showing on the state assessment, funds may be shifted to the support of those subjects where the low scores were identified. The subject areas which are not getting this public exposure may find their financial support reduced. While the potential for better educational decision making is here, there is also the

---

[43]Hencley, Stephen P., "Impediments to Accountability," Administrator's Notebook, Vol. XX, University of Chicago, December, 1971.

potential for poorer educational decision making because of the impact of public exposure of the assessment results on the decision makers.

The foreign languages are not included in the ten subject areas being assessed. The fact that they are omitted means that no assessment is being made of our foreign language skills. Language departments will not be able to cite assessment data as evidence that there is need for greater support of their programs. They have been put into a poorer bargaining position by this omission.

So far as this writer knows, no assessment instrument of this type has been produced to date, but the potential for "shortcut" assessment schemes by publishers is certainly there. Why go to all of the work for an expensive local assessment effort such as San Bernardino's when you can buy a commercial assessment package which is based on the released National Assessment exercises and which will provide the school district with results that may be compared to National Assessment results (norms). The exercises may be given and scored by the local teachers, and the results may be compared by them with the National Assessment results which have been included with the commercial assessment package. If the project is handled "right," a "livewire" superintendent can demonstrate that his district is outscoring the National Assessment results (norms) at a very modest cost.

Before leaving this idea of a commercial assessment package based on released National Assessment exercises, the positive potential of this idea should be mentioned. With honesty and

careful application, this approach could be of considerable value to school districts. First, the objectives being assessed by the exercises included in the instrument need to be carefully identified. If the school district finds that these objectives fit their own objectives adequately, the exercises may be used with confidence. Second, there is no reason why teachers should not be able to conduct and score the exercises accurately. A tape-paced administration could be used, and the directions for scoring could be written in such a way that teachers could follow them with good results. If a district collects data about the performance of its students on National Assessment exercises, these could be very valuable data. This would be worthwhile if these data could be collected at a modest cost. There are a good many "ifs" and pitfalls in this proposal, but it is true that accurate, valid data could be collected in this fashion at a reasonable cost for a school district. Undoubtedly, some districts will find this to be a workable plan for them.

Cost is one of the big problems. It took a great deal of money to develop the highly sophisticated program of National Assessment. Currently, the data being provided by National Assessment are of the highest quality, and the results have great promise for promoting educational improvements. Nevertheless, hundreds of local districts are considering their own assessments, and the question of how to reduce the cost is becoming important. Should they develop their own assessment instruments, or should they use a commercial version of the model? For many districts, it is not possible for them to carry out their own assessment program; hence, for many reasons, the second alternative will be used. Over the coming years, the commercial assessment package will probably become reality. The only other alternative open to poorly financed districts

is to resist the pressure to become involved in local assessment.

One of the potential outcomes of an assessment program is the spinoff of research projects designed to investigate questions of causality that are raised by assessment results. Several places in this paper illustrations have been given where further investigation is needed in order to more accurately interpret results. There are any number of potential doctoral studies in the data coming from National Assessment which need to be made before the results may be confidently used in curriculum work. Currently, one of the barriers to conducting a doctoral study of this type is the security measures which National Assessment of Educational Progress believes that it must take to safeguard its operation. At present, the researcher finds himself being confronted with the need for information which he must have to proceed with his study but which is not made available to him. The results are that it is very difficult to conduct these needed studies. It is suggested here that the policies which govern the use of National Assessment data in these kinds of studies need a thorough review at this time. No one is disputing the need for security in order to assure uncontaminated results, but the issue of overcaution is being raised by this writer. These spinoff kinds of studies have the potential of making an important contribution to the overall effort in assessment, and they should not be killed off by bureaucratic kinds of decisions.

There is a final interesting prospect in this entire assessment business, and that is the increased pace of change in our society within recent decades. Those of us who have been involved in

educational decisions over the last couple of decades know that catching the direction of things is frequently more important than trying to make careful, data-based decisions. By the time we have collected and analyzed our data base, society and technology have gone off and left us. It is hard to fault the soundness of the data collection design of National Assessment, but there is a five-year time lapse from start to finish of a cycle plus the time lag of scoring and data analysis. The question is whether this is the best process for educational decision making in the last quarter of this Century. Are changes in our society coming so fast that long before the data base is established the data are no longer relevant for the decisions for which they were gathered?

Assessment data are badly needed, for they point to the problems and inequalities of our educational system which require attention. The promise of National Assessment and of the adaptations of the model to state and local assessments are great, and the evidence supports moving ahead with this valuable project.

USE OF THE MODEL FOR ACCOUNTABILITY PURPOSES

Assessment is not the same thing as accountability, for accountability places greater emphasis on value judgment than assessment does.[44] It is concerned with the badness or goodness of something. Educational assessment is aimed at improving educational decision making by collecting information concerned with the outcomes of education. Accountability has varying meanings depending on who is writing about it.

A number of approaches have been proposed to make schools more accountable, such as systems approach, management by objectives,

---

[44] Womer, Developing a Large Scale Assessment Program, op. cit., p. 3.

education program auditing, planning-programming-budgeting system, performance contracting, voucher plan, and alternative educational forms.[45] A widely accepted interpretation is that accountability is to determine if the teacher who is assigned the task of educating a group of students is performing that task.[46] On the other hand, a broader interpretation of accountability is that it is a process for determining if the program of a school, district, or state is producing the student achievement expected with regard to the objectives of the program.[47] The first of these interpretations puts the responsibility for individual pupil achievement on the teacher. The second places the responsibility for the output of a program of instruction on the school, district, or state school system.

The first interpretation of accountability has been widely debated and has gained the opposition of many groups including the teachers' organizations. A model of this approach to accountability is given below.

---

[45] Browder, Lesley H., Jr., An Administrator's Handbook on Educational Accountability, American Association of School Administrators, Arlington, Virginia, 1973, pp. 14-21.

[46] Wrightstone, Hogan and Abbott, op. cit., p. 1.

[47] Gronlund, Norman E., (Ed.), Readings in Measurement and Evaluation, Lee J. Cronbach, "Evaluation for Course Improvement," Macmillan Company, New York, 1968, p. 41.

ACCOUNTABILITY MODEL[48]

I. Behavioral Objectives

II. Stated Evaluation Criteria Related to the Objectives

III. Scheduled Materials, Learning Activities, Equipment, Etc.

IV. Teaching Activities

V. Evaluation of Student Performance Based on Behavioral Objectives

Not only does this model stress the output of the teacher in terms of pupil achievement, but it does not provide for the multitude of variables which are found in any teaching situation. The primary concern in the model is, "Are the children learning what the teacher was hired to teach them?" No allowances are made for the many variables such as pupil ability, parental education, or wealth of school district, which may influence the success of the teaching in a given learning situation.

Under this approach to accountability, complete data must be collected for each individual in the population, for here it is necessary to identify the performance of specific individuals with respect to stated objectives since the responsibility for the individual student's performance is to be assigned to the instructor who was charged with teaching for these objectives. Here, it would seem that accountability is taking us back to yearly, mass testing with which some of us are only too familiar. While it is hoped that some of the improvements in test administration such as tape-paced administration introduced by the National Assessment model would be used, it is suspected that because of financial considerations, teachers would be involved in the administration of these

[48]Browder, op. cit., p. 50.

tests with many of the problems concerning results which have been true for this in the past.

The second interpretation of accountability presented here stresses the performance of a given group of students, and it is concerned with what the strengths and weaknesses of a program are, as well as where students are or are not achieving. The National Assessment model may easily be adapted for this type of accountability use. In assessment, the objectives are identified, the achievement level with respect to these objectives in a population is determined through a sampling assessment procedure, and the results are reported in terms of what percentages of the population are achieving the objectives.

The National Assessment model, as it was originally designed, did not provide useable data for determining who was responsible for the individual student either achieving or not achieving the stated objectives. As designed, the model provided information concerning the achievement on the stated objectives of a population or subgroups of that population; hence, it was not possible to identify the results for individuals in these population groups. Also, it was not possible to establish what individual teacher was responsible for the students having either achieved or not achieved the stated objectives.

Now, let us examine the National Assessment model for its application to the evaluation of an entire curriculum for accountability purposes. The model components identified in this paper are: objectives development, exercises development, sampling plan, administration of exercises, scoring and analysis, reporting and

dissemination, and utilization. Under accountability, objectives may be developed with input from a number of sources including the patrons who are paying the bill for education, or an already established set of objectives may be used.[49] Since this is the evaluation of a given curriculum, the objectives should apply to that program if the results are to be valid. Who should select these objectives is an issue in accountability. Shouldn't there be input here from the teachers who are presenting the program?

While there are some problems with respect to how objectives are selected, the components of the National Assessment model are all functional in this adaptation of it for accountability purposes. Measurement specialists contend that objective-referenced exercises which are written for specific learning objectives are better for use in accountability evaluation than the norm-referenced exercises commonly used in standardized tests.[50] Therefore, released objective-referenced, National Assessment exercises may be used for accountability purposes where their objectives are valid for the learning situation which is being evaluated.

In the National Assessment design, the samples were drawn in such a fashion that they represented the population of 9, 13, 17, and 26-35 year olds and subgroups from these populations. Where the model is being used to evaluate the performance of a given population, sampling procedures may be used as they were developed for the National Assessment model. Also, the same professional care which is taken in exercise administration of the National

---

[49]Morrissett, Irving, "Accountability, Needs Assessment and Social Studies," Social Education 37:274, April, 1973.

[50]Wrightstone, op. cit., p. 5.

Assessment model would be desirable in the application of the model for accountability purposes. Scoring and reporting are the same for assessment and accountability, and the same percentage kind of reporting can be used in both situations.

Where accountability is being applied to a total organization such as a school, a district, or a state, the National Assessment model may be used with little or no modification. It was designed to accurately establish what the level of performance on a given set of objectives was in a population, and it can be used to do this for accountability purposes as well as assessment purposes. Likewise, it can assess subgroups of the population and identify specific strengths or weaknesses in the performance of a given subgroup. The model is an excellent instrument for carrying out accountability in this kind of situation.

There are many value questions related to accountability, and while it is not the task of this paper to discuss them, a number of them are cited here. Will tight state accountability structures severely limit creativity and innovation in the schools? As indicated in some of the previous comments, doesn't accountability have the potential of becoming a stulifying state testing program? Will the single-minded pursuit of achieving performance objectives lead to the abuse of children? Couldn't a disproportionate amount of time be expended on the defining of objectives without a commensurate increase in learning? Could not objectives established by politically oriented groups set schools on a course of indoctrination? Currently, in education there is a humanist-behaviorist conflict. Doesn't the accountability movement support a kind of

techno-urban fascism?[51] Could not the accountability movement lead to the teaching of easily quantifiable material and discourage the inclusion of material which is difficult to quantify? Doesn't accountability promote a closed rather than an open educational system? Doesn't accountability present obstacles to the continued development of freedom and autonomy for teachers? This is not an exhaustive list, but these are all disturbing questions. Furthermore, there is considerable evidence in the literature which supports the seriousness of the issues being raised by these questions.[52,53,54,55,56]

SUMMARY

In closing, there are several summary statements which can be made. The National Assessment model is no doubt the best and most comprehensive procedure that has been designed for collecting data for these purposes. The model can and has been successfully adapted for use at the state and local levels, and as states become more active in assessment, there will be many adaptations of it made to fit local needs. Finally, where the concern is with the evaluation

---

[51] Hencley, op. cit.

[52] Gubser, M. M., "Accountability As a Smoke Screen for Political Indoctrination in Arizona," Phi Delta Kappan 55:64-5, September, 1973.

[53] Hand, Harold C., "National Assessment Viewed As the Camel's Nose," Phi Delta Kappan 47:8-13, September, 1965.

[54] House, Ernest R., "The Price of Productivity: Who Pays?" Today's Education 62:65-9, September, 1973.

[55] Landers, Jacob, "Accountability and Progress by Nomenclature: Old Ideas in New Bottles," Phi Delta Kappan 54:539-41, April, 1973.

[56] Ornstein, Allan C. and Harriet Talmage, "The Rhetoric and the Realities of Accountability," Today's Education 62:70-80, September, 1973.

of group performance, the model may be used for accountability purposes. With the current rapid development of state assessment and accountability programs, it is expected that the model will be widely used for these purposes in the coming years.

BIBLIOGRAPHY

Bayless, David L., Ralph E. Folsom, and Louise H. Lewis, "Sample Design for Assessing Montgomery County Public Schools 13- and 17-Year-Old Pupils Using the NAEP Model," National Assessment of Educational Progress and Educational Commission of the States, Denver, Colorado, January, 1973.

Bettinghaus, Erwin P. and Gerald R. Miller, "A Dissemination System State Accountability Programs, Part I," Cooperative Accountability Project, Denver, Colorado, June, 1973.

Bonney, Lewis A., "Accountability at the Local Level," Unpublished Paper, San Bernardino City Unified School District, San Bernardino, California.

"Application of the National Assessment of Educational Progress Philosophy in San Bernardino City Unified School District," Unpublished Paper, San Bernardino City Unified School District, San Bernardino, California.

Brandt, Ronald, Associate Superintendent for Instruction, Report on Assessment Results to Board of Education, Lincoln Public Schools, Lincoln, Nebraska, Spring, 1973.

Browder, Lesley H., Jr., An Administrator's Handbook on Educational Accountability, American Association of School Administrators, Arlington, Virginia, 1973.

Campbell, Vincent N. and Daryl G. Nichols, "National Assessment of Citizenship Education," Social Education 32:279-81, June, 1969.

Campbell, Vincent N., Manford J. Ferris, and Daryl G. Nichols, National Assessment Report 6, 1969-1970 Citizenship: Group Results for Sex, Region, and Size of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, July, 1971.

Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, November, 1970.

Campbell, Vincent N., et al, Citizenship Objectives for 1974-75 Assessment, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, 1972.

Conaway, Larry E., "Some Implications of the National Assessment Model and Data for State and Local Education," Paper Presented at the 1973 Annual Meeting of the AERA, New Orleans, Louisiana, February 26, 1973.

Fairley, Richard L., "Accountability's New Tool," American Education 5:33-5, June, 1972.

Finley, Carmen J., "National Assessment: Reports and Implications for School Districts," National Elementary School Principals 50:25-32, January, 1971.

Finley, Carmen J. and Frances S. Berdie, The National Assessment Approach to Exercise Development, National Assessment of Educational Progress, Ann Arbor, Michigan, 1970.

Gadway, Charles J. (Ed.), Reading and Literature: General Information Yearbook, Education Commission of the States, Report 02-GIY, National Assessment of Educational Progress, Denver, Colorado, May, 1972.

Glass, Gene V., "The Many Faces of Educational Accountability," Phi Delta Kappan 53:636-9, June, 1972.

Gronlund, Norman E., (Ed.), Readings in Measurement and Evaluation, Lee J. Cronbach, "Evaluation for Course Improvement," Macmillan Company, New York, 1968.

Gubser, M. M., "Accountability As a Smoke Screen for Political Indoctrination in Arizona," Phi Delta Kappan 55:64-5, September, 1973.

Hand, Harold C., "National Assessment Viewed As the Camel's Nose," Phi Delta Kappan 47:8-13, September, 1965.

Harrison, Charles H., "Are We Educating for Tomorrow?" Scholastic Teacher, September 21, 1970.

Helper, John W., An Assessment of Learner Needs in Colorado, Colorado Department of Education, Denver, Colorado, 1972.

Hencley, Stephen P., "Impediments to Accountability," Administrator's Notebook, Vol. XX, University of Chicago, December, 1971.

House, Ernest R., "The Price of Productivity: Who Pays?" Today's Education 62:65-9, September, 1973.

Kurfman, Dana G., "A National Assessment of Social Studies Education," Social Education 31:209-11, March, 1967.

Evaluating Geographic Learning, 1970 Yearbook, National Council for the Social Studies, Washington, 1970.

Landers, Jacob, "Accountability and Progress by Nomenclature: Old Ideas in New Bottles," Phi Delta Kappan 54:539-41, April, 1973.

Maine Assessment of Educational Progress: Methodology (Report 5), Department of Educational and Cultural Services, Augusta, Maine, 1972.

Maine Assessment of Educational Progress (Report 2):Results Report I, Citizenship and Writing, Department of Educational and Cultural Services, Augusta, Maine, 1972.

Morrissett, Irving, "Accountability, Needs Assessment and Social Studies," Social Education 37:271-79, April, 1973.

Norris, Eleanor L. (Ed.), Citizenship Objectives, Committee on Assessing the Progress of Education, Ann Arbor, Michigan, 1969.

Norris, Eleanor L. and Barbara Goodwin (Eds.), Social Studies Objectives, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, 1970.

Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size and Type of Community, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, May, 1972.

Norris, Eleanor L., et al, Report 1, 1969-1970 Science: National Results and Illustrations of Group Comparisons, J. R. Chrony and D. G. Horvitz, "Structure of Sampling and Weighting," Appendix C, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, July, 1970.

Ornstein, Allan C. and Harriet Talmage, "The Rhetoric and the Realities of Accountability," Today's Education 62:70-80, September, 1973.

Report on the Assessment of Reading Skills of Connecticut Public School Students, Institute for the Study of Inquiring Systems, Philadelphia, Pennsylvania, and Department of Education, Hartford, Connecticut, 1972.

Scriven, Michael, "Prose and Cons About Goal-Free Evaluation," Evaluation Comment 3:1-4, December, 1972.

Sixth-Grade Mathematics: A Needs Assessment Report, Texas Education Agency, Austin, Texas, 1972.

Special Curriculum Task Force, "Report on Student Performance," Office of Instructional Services and Research and Development Office, San Bernardino City Unified School District, San Bernardino, California, June, 1972.

Stake, Robert E., "National Assessment," Proceedings of the 1970 Invitational Conference on Testing Problems--The Promise and Perils of Educational Information Systems, Educational Testing Service, Princeton, New Jersey, 1971.

State Educational Assessment Programs, 1973 Revision, Joan S. Beers and Paul B. Campbell, "Statewide Educational Assessment," Educational Testing Service, Princeton, New Jersey, 1973.

Von Ende, Joe, "Layman's View," Compact 6:23, February, 1972.

"Weekly Focus," Lincoln Public Schools, Lincoln, Nebraska, February 12-19, 1973.

Womer, Frank B., Developing a Large Scale Assessment Program, Cooperative Accountability Project, Denver, Colorado, 1973.

What Is National Assessment? National Assessment of Educational Progress, Denver, Colorado, 1970.

Womer, Frank B. and Marjorie M. Mastie, "How Will National Assessment Change American Education?" Phi Delta Kappan 53:118-20, October, 1971.

Wrightstone, J. Wayne, Thomas P. Hogan, and Muriel M. Abbott, "Accountability in Education and Measurement Problems," Test Service Notebook 33, Harcourt, Brace, Javanovich, Inc., New York.